DOCUMENT RESUME

ED 138 650 TM 006 315

AUTHOR Fox, G. Thomas, Jr.; And Others
TITLE Reflections Upon the Use of Time-Series Design and Analysis for Program Evaluation.
PUB DATE [Apr 77]
NOTE 65p.; Paper presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$3.50 Plus Postage.
DESCRIPTORS Data Analysis; *Evaluation Methods; *Program Evaluation; Statistical Analysis; Student Teachers; Teacher Education; Time
IDENTIFIERS Teacher Corps; *Time Series Analysis

ABSTRACT

An alternative evaluation procedure is critiqued that utilizes both a time-series design and a time-series data analytic procedure. The bases for this critique are the authors' experiences and reflections upon their involvement in a recently completed evaluation project that used a time-series approach. The study was a sixteen month evaluation of a one month training event and its impact on teacher trainees during the following school year. The study was sponsored by Teacher Corps, Washington. The one month training session was the Teacher Corps Member Training Institute (CMTI) held in July, 1975. The critique of the time-series design and the supporting data analytic procedure focuses upon their application to program evaluation. Three questions are posed and answered in this critique: (1) What is time-series description and time-series analysis?, (2) What can time-series design and analysis tell the evaluator? (i.e., Are there certain kinds of evaluation questions that are especially applicable to a time-series design and analysis?), and (3) How can one make time-series design and time-series analysis work in an evaluative setting? (i.e., some practical considerations and limitations when applying the time-series design and analysis to program evaluation). Where applicable, examples are used from the CMTI evaluation study. Time-series is suggested to be an alternative evaluation design and data analytic procedure that can be especially applicable to evaluators who wish to illuminate the process and the impact of an educational program as well as measure its effect in terms of intended outcomes. (Author/RC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED138650

Reflections Upon the Use of Time-Series Design and Analysis for Program Evaluation

G. Thomas Fox, Jr. University of Wisconsin-Madison

Leo W. Anglin, Kent State University

Thomas A. Romberg, University of Wisconsin-Madison

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY

Thomas A. Romberg

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Paper delivered at 1977 AERA Conference, April 6, 1977.
This is a draft. We request that you do not reproduce this paper without permission of the authors.

TM006 315

ABSTRACT

Recent experiences in program evaluation have demonstrated the limitations of many standard evaluation procedures. The purpose of this paper is to critique an alternative evaluation procedure that utilizes both a time-series design and a time-series data analytic procedure. The bases for this critique are the authors' experiences and reflections upon their involvement in a recently completed evaluation project that used a time-series approach.

The study was a sixteen month evaluation of a one month training event and its impact on teacher trainees during the following school year. The study was sponsored by Teacher Corps, Washington. The one month training session was the Teacher Corps Corps Member Training Institute (CMTI) held in July, 1975.

The critique of the time-series design and the supporting data analytic procedure focuses upon their application to program evaluation. Three questions are posed and answered in this critique: (1) What is time-series description and time-series analysis?, (2) What can time-series design and analysis tell the evaluator? (i.e., Are there certain kinds of evaluation questions that are especially applicable to a time-series design and analysis?), and (3) How can one make time-series design and time-series analysis work in an evaluative setting? (i.e., some practical considerations and limitations when applying the time-series design and analysis to program evaluation). Where applicable, examples are used from the CMTI evaluation study.

The paper concludes that, even considering the unique limitations to the time-series approach, it does offer a viable alternative to evaluation of educational programs. The nature of the information supplied and the inferences supported by a time-series design and data analytic process can be extremely appropriate to illuminating learning processes, instructional effects and the research of concomitant variation, for more understanding of unexpected and unanticipated consequences as well as a greater attention to the unintended intervention procedures that are a major feature of many educational settings. Time-series, then, is suggested to be an alternative evaluation design and data analytic procedure that can be especially applicable to evaluators who wish to illuminate the process and the impact of an educational program as well as measure its effect in terms of intended outcomes.

ACKNOWLEDGMENTS

The authors would like to express their appreciation to Linda Lee Weiler, who was both secretary and administrative assistant to this project, and to Ann Campoll, who typed this draft of the paper. John Griffith is the artist for the figures in this paper. We would also like to acknowledge the importance of our dialogue and interaction with other investigators of this evaluation study: Drs. Carl Grant, Tom Popkewitz, Bob Tabachnick and Gary Wehlage.

Introduction

Program evaluation is the art of the possible in an arena not normally chosen, controlled or designed by the evaluators. Past efforts in educational program evaluation have demonstrated the limitations of what has come to be known as standard evaluation procedures and the corresponding need for alternative forms of evaluation (see Fox, 1976; Ross and Cronbach, 1977). Time-series has recently been suggested as one alternative design and analytic procedure for educational program evaluation (see Glass, Willson and Gottman, 1975; Campbell, 1975; Romberg and Fox, 1976). Unfortunately, most examples of time-series designs and data analyses come from either experimental research settings or in evaluations of non-educational settings (e.g. Campbell, 1975; Glass et al, 1975).

The authors of this paper have recently applied a time-series design and time-series analysis to an educational program evaluation. The program we evaluated was a one month training event in teacher education that had many of the characteristics that define the uniquely difficult arenas of educational programs. The training event was to be given to all participants in the program, thus leaving no chance for control groups or for randomization from a larger population. It was a time-bounded single event, thus leaving no chance to go back in time to the event or setting to check on certain descriptions or interpretations. It was a definite innovation where no one was certain what may be an outcome (either intended or unintended, nor even what may be appropriate or

inappropriate interpretations of outcomes). It was costly in time, effort and resources which meant that there were specific decisions that the client agency needed to make about the feasibility and desirability of supporting similar events. The planners, organizers and instructors of the event were heavily involved in and committed to their intent and instructional design, thus making any relationship between them and the evaluation effort sensitive. The political arrangements between supporting agency, instructional leaders, instructors, participants and evaluators were always ambiguous, thus making coordination, negotiation and communication always a necessary but tricky process. In short, it presented a normal, everyday evaluation problem for which there are no ready-made solutions.

We took, as our solution, time-series design and data analysis. Before reflecting upon our experience in applying time-series analysis to program evaluation, it may be helpful to provide some background to the study.

Background: The Evaluation Problem

The general problem of our study was to evaluate the impact of a one month training event upon student teacher participants. The event and the impact study were sponsored by Teacher Corps, a federal program funded since 1965 to improve the schooling environment of children from low-income communities through professional development applied by local projects. (Teacher Corps funds about sixty projects a year for a two year period.) Starting in June 1975, an original Teacher Corps

emphasis on pre-service teacher education was changed by Congress to include training of experienced teachers. In operational terms this meant that each Teacher Corps project would have four pre-service student teacher candidates instead of the previous thirty-five.[1] The added focus on retraining experienced teachers led Teacher Corps to anticipate some consequences for the pre-service student teachers. Some of the anticipated consequences were that student teachers would have less attention, less status, less interaction among themselves and less cultural representation within each project. One response to these anticipated consequences was to plan a new Teacher Corps undertaking: The 1975 Corps Member Training Institute.

The 1975 Corps Member Training Institute (CMTI) was to be a one month training session for approximately two hundred and twenty-five student teachers from fifty-four different Teacher Corps projects throughout the United States and Micronesia. Included as purposes of the CMTI training event were: (i) to provide a rich multi-cultural interaction experience for participants, (ii) to develop an esprit de corps among participants, (iii) to provide a background for participants to develop perspectives towards schools as organizations, and (iv) to introduce the participants to specific models of teaching.

Since the CMTI was both a new undertaking for Teacher Corps and also reflected an innovative feature to teacher education (training in organizational theory), an impact study was created as part of the instructional contract.[2] The responsibilities given this impact study

were to demonstrate the nature of the training event, to determine the impact of the event upon participants, to provide policy relevant information for possible future replication of the event and to communicate the event and its impact to the teacher education profession at large.

One of the first responses to these responsibilities was to create a multi-disciplinary evaluation team of teacher educators.[3] This team participated in the description, interpretation, analyses and the writing of the study's reports.

Initial decisions made by the impact study team help clarify the way in which we interpreted our responsibilities. First, we considered the CMTI to be a one month event within student teacher participants' two year teacher education program. This meant that to answer the basic question, "What kind of effect did the CMTI event have upon the participants?", we needed to look beyond the time boundaries of CMTI to determine its impact. Second, the study team decided to look at teacher education as a socialization process. This meant that determining impact was interpreted in a wider context than the stated instructional objectives of CMTI. The chosen context was the impact of the training event upon the participants' development of professional perspectives towards teaching. Third, we decided that interpretations and conclusions of impact must include a synthesis of collected data.[4] This meant that the relationship between quantitative and qualitative data was not considered to be one of simple effect and cause or of dependent and independent variables. Instead observational descriptions of what happened, participant interview responses, survey results, instructional

intentions and instructors' underlying assumptions were to be related when interpreting the impact of the training event upon participants. The team also decided to use photography in evaluation, not only for record keeping but in an interpretive and analytic role.[5]

Generally, then, the team approached the problem of determining the impact of the Teacher Corps, 1975 Corps Member Training Institute as a real challenge to capture, interpret and illuminate the effects of an anticipated dynamic event within the professional life of the participants.

The Application of a Time-Series Design and Analysis to the Problem

Our response to this challenge to capture, interpret and illuminate the effects of an anticipated dynamic event was to apply a single-group multiple intervention time-series design as described by Campbell and Stanley (1963), suggested by Campbell (1976) and refined by Glass, Willson and Gottman (1975). Major features of the application of this time-series design were:

(1) multiple readings across time on a dependent variable measured by survey instruments,

(2) a thorough description of specific events along the same time continuum,

(3) the use of both descriptive statistics and the ARIMA inferential time-series procedures to analyze the survey data and

(4) the use of a continuous, interactive team process that periodically interpreted meaningful fluctuations of the survey results

and related these fluctuations with participants' expectations, actions, responses and assumptions that were being captured by the event descriptions.

The following describes our application of the single-group multiple intervention time-series design and the ARIMA data analysis in more detail.

Study Design

The major purpose of the study was to demonstrate the impact of the event upon participants' development of professional perspectives. To achieve this purpose, the study was separated into two phases: the event phase and the follow-up phase. The tasks of the event phase were: to describe the event, to provide information about interns' development of professional perspectives during the event, and to suggest the most interesting directions to continue to pursue in demonstrating the impact of the event. The tasks of the follow-up phase were: to answer the questions raised by the description phase, to describe the various local site experiences that participants had in the school year following the event, to interpret the residual effects of the event within the context of these local site experiences, and to develop the most viable interpretations of the impact of that event upon the continuing development of participants' professional perspectives.

Five perspectives on teaching were identified and survey instruments were designed to elicit the rank and value of these five perspectives from participants. These surveys were administered eighteen different times during the study: eight times during the one month training

event (at the beginning and end of each week) and ten times during the follow-up phases. In addition to showing the time of the survey administrations, Figure 1 points out three major limitations to our application of the time-series design. There were only eighteen data points for the analysis, there was no preliminary data base collected before the event and there was a large time lag between the end of the event and the beginning of the survey administrations for the follow-up phase. There were reasons for each of these limitations that will be discussed later.

In intensive field descriptions and interpretations, the study certainly more than met the suggestion by Glass et al (1975) for context information beyond the repeated observation measures. Over two hundred interviews and observations were reported during the one month training event. Photographs were employed to help analyze participants' interactions during the event. All instructional materials used in the event were analyzed from three different curriculum perspectives. Sixteen two-day on-site visits (to eleven sites) were conducted during the follow-up phase. Group interviews were held with representatives from nearly all of the fifty-three sites at the end of the follow-up phase.

Given the limited population of the participants (about 225), the number of administrations of the survey and the complexity of the settings after the shared event (participants spent the remainder of their training in one of fifty-three different sites throughout the United States), sampling was (like in most program evaluation studies) a problem

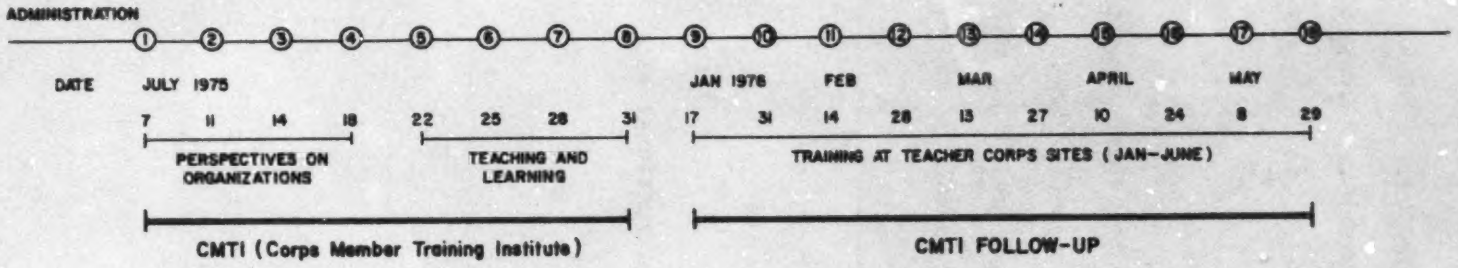

ADMINISTRATION
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
DATE
JULY 1975
JAN 1976
FEB
MAR
APRIL
MAY
7 11 14 18 22 25 28 31 17 31 14 28 13 27 10 24 8 29
PERSPECTIVES ON ORGANIZATIONS
TEACHING AND LEARNING
TRAINING AT TEACHER CORPS SITES (JAN–JUNE)
CMTI (Corps Member Training Institute)
CMTI FOLLOW-UP

Figure 1. Schedule for Survey Administrations

to this study. Our solution was to use matrix sampling for the survey administrations. During the training event, this included a random sampling depending upon the type of questionnaire previously answered. For the follow-up phase, the matrix sampling was based upon five training strategies being used by each site.

Data Analysis

Two of the four surveys used were: participants' views on what teachers do (called Teacher Task Survey) and participants' views of what student teachers should learn (called Intern Task Survey). These surveys consisted of five sets of items, each item an example of one of the five perspectives on teaching (see Figure 2). Respondents were

Figure 2. Sample of one set of items for Intern Task Survey.

IV. My tasks as an INTERN are to learn how to:	Rank Order	Value-to-me (high ○○○○○ low)
a. Extend school programs into the community. . .	_____	○○○○○
b. Continually evaluate the implications of schooling as an organizational enterprise. . .	_____	○○○○○
c. Design basic skills materials which are particularly appropriate for low-income communities	_____	○○○○○
d. Teach children reading comprehension skills. .	_____	○○○○○
e. Be open, warm and sympathetic to my students..	_____	○○○○○

asked, first, to rank order each set of five items and, second, to determine the value of each item. Thus, for each survey on each perspective on teaching, two scores were generated: the mean rank score

and the mean value score for each administration. For example, the mean rank of teaching as technique as a task of an intern was reported for each of the eighteen administrations. In addition to the two mean scores, variances and 95% confidence intervals were computed for the five professional perspectives in each survey.

An ARIMA time-series analysis described by Glass et al was conducted on the mean rank and mean value for each perspective on teaching for these two surveys. Since no baseline data was collected before the actual training event, the two "interventions" for the time-series analysis became the curriculum break between the second and third weeks of the training event, and the time between the end of the training event and the beginning of the follow-up surveys (i.e. between administrations 4 and 5 and between administrations 8 and 9 respectively, see Figure 1).

As an example, the ARIMA time-series analysis of the Teacher Task Survey on teaching as technique showed the following results. As portrayed by Figure 3 and further explicated by Tables 1 and 2, the mean rank scores for teaching as technique on the teacher task survey fluctuated during administrations 1-4, leveled off high during the rest of the training event (administrations 5-8) and fluctuated again during the follow-up phase (administrations 9-18). Using the ARIMA time-series analysis, however, it is interesting to note that a stationary model for the rank means (see Tables 3 and 4) did not yield a significant *t*-statistic for change in level or change in direction for either the first intervention (the break in the training event) or the second intervention (the time between the end of the event and the beginning of the follow-up

Figure 3 Mean Rank and Value for Teaching as Technique (Teacher Tasks)

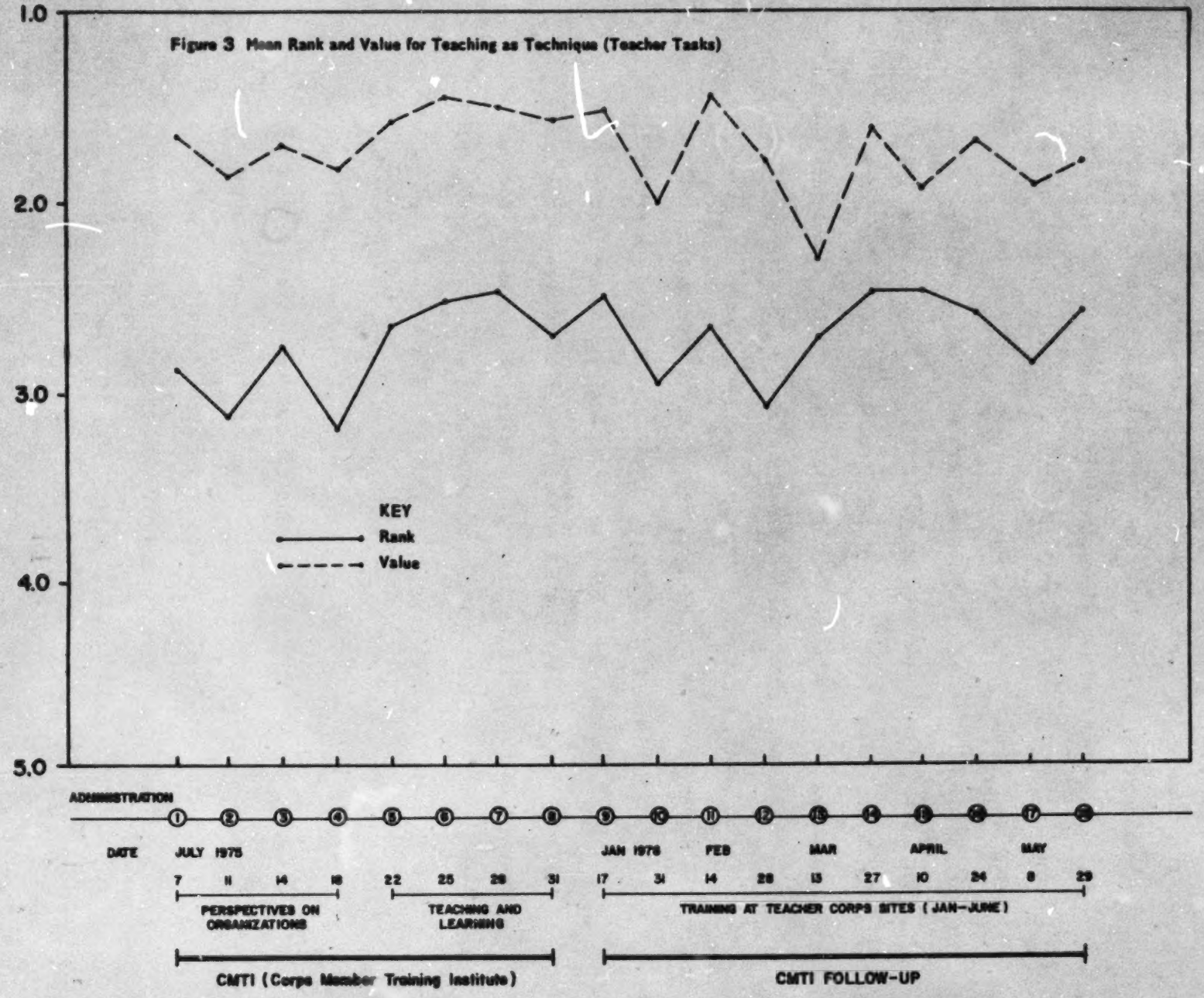


Table 1 CMTI (Teacher Tasks)

Summary Statistics for Teaching as Technique

		RANK			VALUE		
Administration	Number	Mean	Variance	Confidence Interval	Mean	Variance	Confidence Interval
1	37	2.87	.38	2.67 - 3.08	1.65	.28	1.47 - 1.82
2	38	3.12	.54	2.88 - 3.36	1.85	.34	1.65 - 2.04
3	44	2.74	.62	2.50 - 2.98	1.69	.46	1.48 - 1.90
4	40	3.20	.57	2.95 - 3.41	1.81	.53	1.57 - 2.04
5	44	2.63	.52	2.41 - 2.85	1.57	.39	1.38 - 1.76
6	38	2.51	.54	2.27 - 2.75	1.45	.25	1.28 - 1.61
7	37	2.47	.45	2.25 - 2.69	1.50	.26	1.33 - 1.67
8	36	2.69	.52	2.45 - 2.94	1.56	.26	1.39 - 1.73

Table 2 CMTI Follow-up (Teacher Tasks)

Summary Statistics for Teaching as Technique

		RANK			VALUE		
Administration	Number	Mean	Variance	Confidence Interval	Mean	Variance	Confidence Interval
9	9	2.49	.44	1.98 - 3.00	1.51	.19	1.18 - 1.85
10	19	2.94	.65	2.55 - 3.33	2.00	.64	1.61 - 2.39
11	5	2.64	.91	1.46 - 3.82	1.44	.11	1.03 - 1.85
12	12	3.07	.67	2.55 - 3.59	1.77	.46	1.33 - 2.19
13	8	2.70	.74	1.98 - 3.42	2.30	.12	1.40 - 3.20
14	18	2.47	.39	2.15 - 2.78	1.60	.31	1.32 - 1.88
15	23	2.47	.87	2.07 - 2.88	1.92	.48	1.62 - 2.22
16	18	2.57	.80	2.12 - 3.01	1.67	.30	1.39 - 1.94
17	8	2.83	.37	2.32 - 3.34	1.90	.45	1.34 - 2.46
18	22	2.56	.68	2.20 - 2.93	1.77	.51	1.46 - 2.09

Table 3 Teaching as Technique (Teacher Tasks - Rank and Value)

CMTI Time-Series Analysis Data

$N_1 = 4$, $N_2 = 4$, Degrees of Freedom for this Design = 4

Model Identification: Order of Autoregression is 1, Order of Differencing is 0, Order of Moving Average is 0

	Phi or Theta	Error Variance	Level	T-Stat	Level Change	T-Stat	Drift	T-Stat	Drift Change	T-Stat
Rank	-.72	.022	2.89	21.84	-.41	-2.24	.04	.73	-.06	-.87
	Chi-Square Statistics = 1.54 (For Residuals of Difference Order 0, Lags 1 to 2)									
Value	-.64	.005	1.69	26.36	-.28	-3.22*	.03	1.10	-.03	-.82
	Chi-Square Statistics = .75 (For Residuals of Difference Order 0, Lags 1 to 2)									

Table 4 Teaching as Technique (Teacher Tasks - Rank and Value)

CMTI Follow-up Time-Series Analysis Data

$N_1 = 8$, $N_2 = 10$, Degrees of Freedom for this Design = 14

Model Identification: Order of Autoregression is 1, Order of Differencing is 0, Order of Moving Average is 0

	Phi or Theta	Error Variance	Level	T-Stat	Level Change	T-Stat	Drift	T-Stat	Drift Change	T-Stat
Rank	-.40	.041	3.14	25.77	.32	2.08	-.08	-3.36	.06	2.11
	Chi-Square Statistics = 1.64 (For Residuals of Difference Order 0, Lags 1 to 4)									
Value	-.52	.033	1.86	18.25	.28	2.17*	-.05	-2.47	.06	2.59*
	Chi-Square Statistics = 2.84 (For Residuals of Difference Order 0, Lags 1 to 4)									

*p < .05
**p < .01

surveys). Thus, the time-series analysis suggests that the fluctuations noted in comparative rankings were not significantly affected by either "intervention."

The time-series analysis on the mean value scores, however, does show significant changes after each intervention. For the first "intervention", time-series analysis conducted with a stationary model yielded a significant t-statistic for change in level of mean value scores and a nonsignificant t-statistic for change in drift (see Table 3). It also yielded significant t-statistics for change in level and change in drift following the second "intervention" (see Table 4). Thus, the time-series analysis suggests that the fluctuations of the comparative rankings were not affected by either intervention but that the value of teaching as technique was effected by both interventions. The "intervention " during the training event increased the value level and the intervention after the training event decreased both the level and the drift. The provocativeness and crankiness of these results should lead us well into the discussion of how time-series design and data analysis contributed to the purposes of the program evaluation.

How Time-Series Design and Data Analysis Was Integrated Into The Program Evaluation

The most salient feature of our application of time-series design and data analysis was the periodic occurence of interpretations of the event and its possible impact. From the beginning of the planning phase of the study, during the training event itself, immediately following

the training event, during the six months of the follow-up phase and through the final round-up of the data and the writing of the final reports, what participants were doing and the impact of these actions upon their professional perspectives were continually being interpreted by members of the evaluation team.

In the description of what participants were doing and in the resultant interpretations of the impact of these actions upon participants, the importance of interaction and dialogue between different members of the study team cannot be overemphasized. Interpretation of the qualitative and quantitative data describing the event and its impact required technical expertise from qualitative investigators and quantitative investigators. Placing meaningful significance upon these interpretations, however, required dialogue between the investigators to share their interpretations of the evidence. Since it was necessary to understand one another, efforts were made to include all team members in both quantitative and qualitative data gathering and interpretation. Qualitative investigators, for example, had major roles in the conceptualization and the writing of the survey instruments; quantitative investigators took part in the on-site visits, interviews and observations. Care was taken to have all members take part in the interpretive sessions since at no time (as will be suggested) was the quantitative and qualitative data self-evident. Thus, the general interpretive process was a team dialogue where strong, forceful interpretations from multiple perspectives were crucial to an evolving sense of both process and impact.

From this perspective, time-series design and analysis was one component of the evaluation study. To understand how time-series design

and analysis was integrated into the program evaluation, it is necessary to discuss three important features of the evaluation process we used. These three features are our choice of underlying assumptions and theoretical perspective, the periodic generation, review and regeneration of hypotheses and the political affiliations and negotiations that were formed both by our intentions and those of the parent agency, instructors and participants. The following discusses these features of our study and shows how time-series design and analysis contributed to these features.

Choice of Underlying Assumptions

The conceptual basis for the study was the underlying assumption that the process of becoming a teacher is a dynamic socialization process (see Popkewitz, Technical Report no. 18; Fox et al, 1976). This meant that we interpreted the one month training event to be imbedded in a larger context of training experiences, that the anticipations, expectations and assumptions of the participants would play a major role in what happened (and the effects of what happened), that the impact of these experiences may be continually changing participants' perspectives of what teaching was but that the accumulation of these periodic perspectives may be tending towards a personal construct of what teaching means.

It was the conceptual framework of teacher development being a socialization process that determined our choice of a time-series design and analysis. Time-series allowed for a variety of intervention effects (e.g. delayed, evolutionary or temporary) as well as for a flexible interpretation of intervention both during and after the training event. The assumptions of time-series analysis, that is, that there exists a dynamic

process that is being intervened upon, that effect of this intervention requires multiple observations across time, that any observation at time t is dependent upon the previous history of observations and that a series of observations may be dynamic and non-stationary were consistent with our chosen theoretical framework.

It was, in fact, the assumption that the process of being socialized into the teaching profession would be dynamic that determined our adoption of a continuous dialectical mode of inquiry where interpretations would necessarily be periodically shared, reviewed and revised. This conceptual framework determined what it was we were looking for: what participants were doing, changes in perspective, critical incidents, participants' views of what was happening, the underlying assumptions of the instructors, interactions between participants. The framework also determined meanings to our findings. Instructional objectives of the training event, for example, were interpreted as intentions rather than outcomes, the instructional materials and process used were interpreted just as much as public announcements of instructors' underlying assumptions as they were interpreted as means to reaching stated objectives. In short, our conceptualization of the event within the theoretical framework of a dynamic socialization process, helped determine both the methodologies we chose to capture information and the interpretations we gave to the data. Time-series design and time-series data analysis were not only consistent with this chosen framework, they would help illuminate what was most meaningful from this perspective. The following describes in some detail how this illumination occurred.

Periodic Generation of Hypotheses

A basic feature of interpretation in this study was the generation, review and re-generation of hypotheses. Each hypothesis became a shared statement about (a) what we considered to be important, (b) what we did not know, and (c) what we expected to happen. Naturally, these hypotheses were formed before the event, but then, as we began to understand what participants were doing and the context of what was happening, our hypotheses began to change as new features became important, surprises occurred and what we thought we knew was challenged. As an example, one hypothesis stated before the training event began was:

> (i) During the event, participants' perspectives of what teachers should do would not differ from their perspectives of what student teachers should learn.

This hypothesis showed our focus upon participants' perspectives towards teaching. It reflected our anticipation that disenchantment may eventually occur, between participants and teachers, but that we also anticipated that during this first month of training there would be no impact since there was no contact with either teachers or the realities of classrooms.[6]

During the event, although interviews with participants suggested that they were much more anxious about what they would be doing in classrooms and how they would get along with both the students and teachers than had been expected, neither the interviews nor our observations challenged this hypothesis. The survey results, however, did challenge our hypothesis. After two weeks time, there seemed to be a definite

split between (a) what participants felt teachers do (e.g. Teacher Task Survey) and (b) what they felt they themselves should be learning (Intern Task Survey). These differences were especially pronounced in the perspectives we categorized as teaching as technique (see Figure 4) and in teaching as community involvement (see Figure 5). These results from the time-series design and the descriptive statistical analyses immediately following the event, suggested revisions in the original hypotheses to reflect our findings and our new concerns. At the beginning of the follow-up phase our hypothesis became:

> (ii) Participants' differentiation between teachers and themselves would become even more pronounced as they began to work with teachers.

Thus, we decided to look more closely at why there was a dramatic change in perspective during the event. Although the interactive process and underlying assumptions that described the first two weeks, followed by a reflective three day period and then a structured process with a very different set of assumptions, seemed to explain some of the dramatic findings shown in Figures 4 and 5, we were not certain of our explanation. For example, we did not have much information on what occurred during the three day break, the time when the shift took place (between administrations 4 and 5). This version of the hypothesis, however, still reflected our assumption from a socialization perspective that it would be the work (what participants <u>did</u> in the classroom) and the nature of that work that would be most challenging to the perspectives participants held towards teaching.

Fig. 4 Summary of Mean Rank of Teaching as Technique

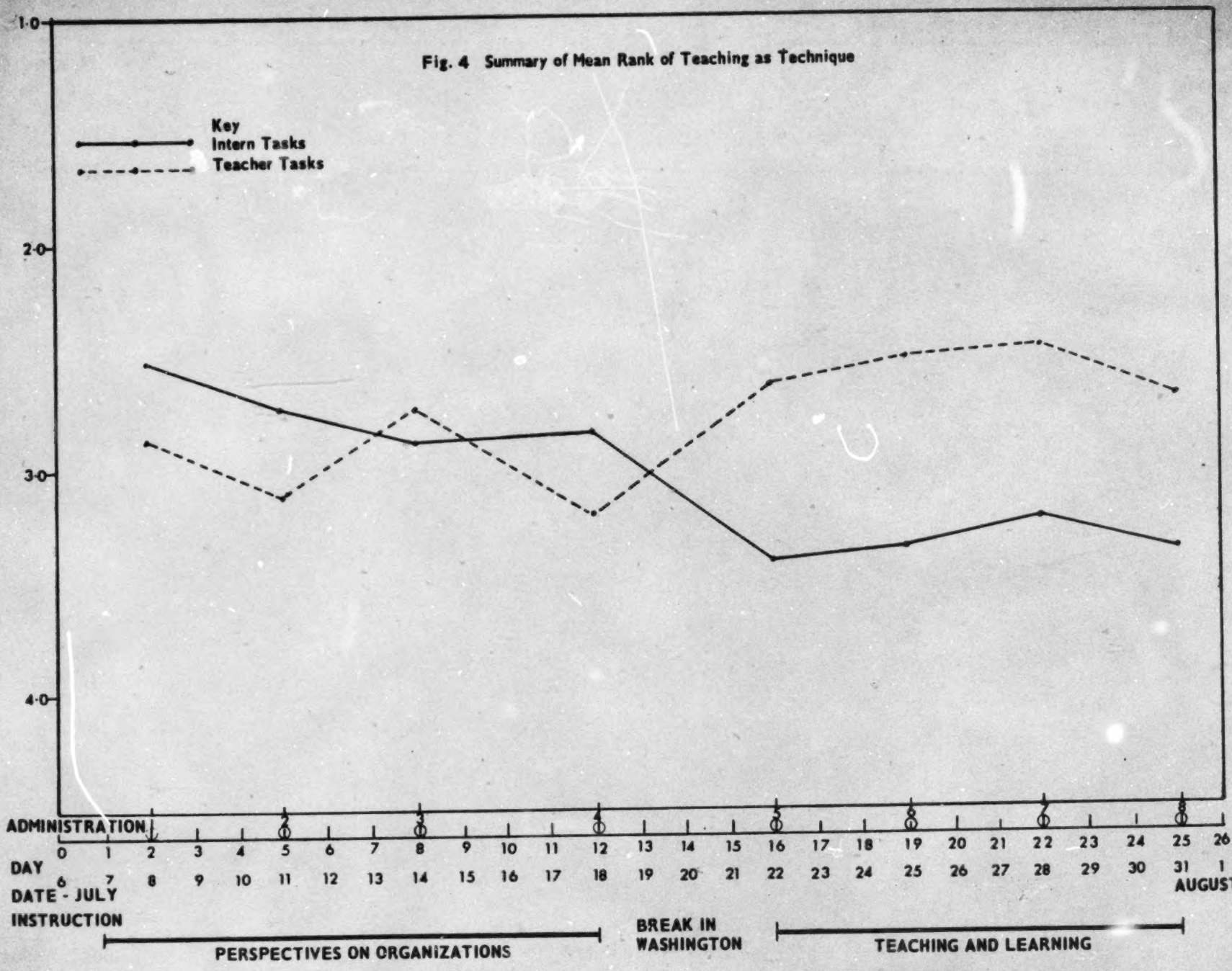
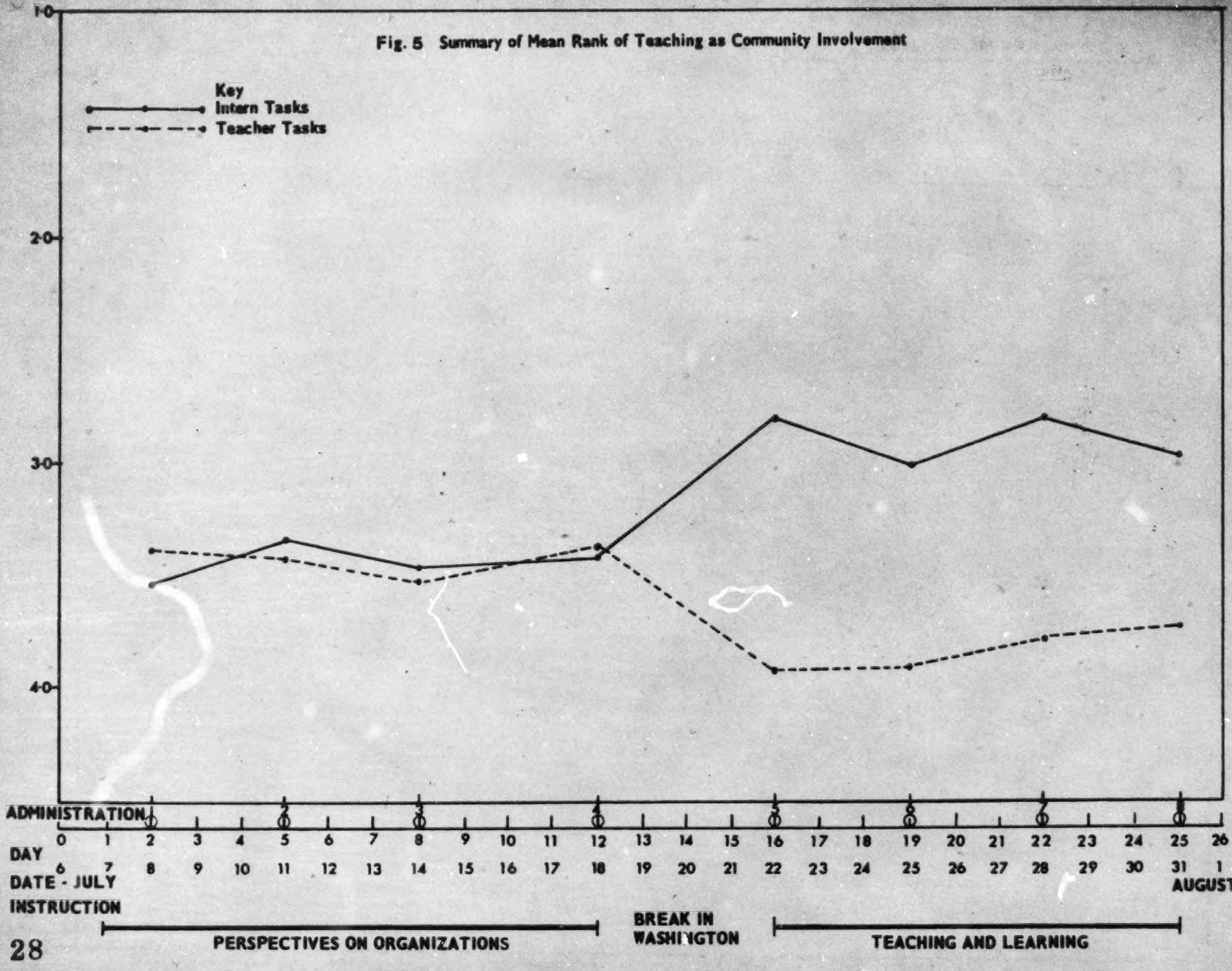


Fig. 5 Summary of Mean Rank of Teaching as Community Involvement


Key
Intern Tasks
Teacher Tasks
1·0
2·0
3·0
4·0
ADMINISTRATION
DAY
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
DATE - JULY
6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 1
AUGUST
INSTRUCTION
PERSPECTIVES ON ORGANIZATIONS
BREAK IN WASHINGTON
TEACHING AND LEARNING

In our discussions with participants and our observation of their actions in classrooms, however, two things began to challenge our new hypothesis (ii). One was that, although some participants verbalized a difference between themselves and their teachers, many did not and, furthermore, in terms of classroom action, about all student teachers looked to be doing the same as teachers. This qualitative finding was later reflected in the time-series results during the follow-up phase. As Figure 6 shows, the original differentiation noted in Figure 4 dissipates. Second was the evidence, again from our discussion with participants, that technique and student personal development did not seem to be viewed by participants as opposite ends of a continuum as we had originally interpreted from the results shown in Figure 7. In these interviews, it became apparent that technique and student personal development were not mutually exclusive perspectives on teaching. Thus, about mid-way through the follow-up phase of the study, we began to hypothesize the following:

(iii) The relationships being developed between teaching as technique and teaching as student personal development are crucial factors in participants' development of perspectives towards teaching.

Now our concern was not so much about participants' differentiating themselves from teachers but was focused more upon the meanings behind the dynamic change (noted in Figure 7) and the ambiguities we were findings about their views on teaching during the follow-up phase (see Figure 3). The more we discussed teaching with participants, the

Figure 6 Summary Comparison of Mean Rank of Teaching as Technique (Intern and Teacher Tasks)

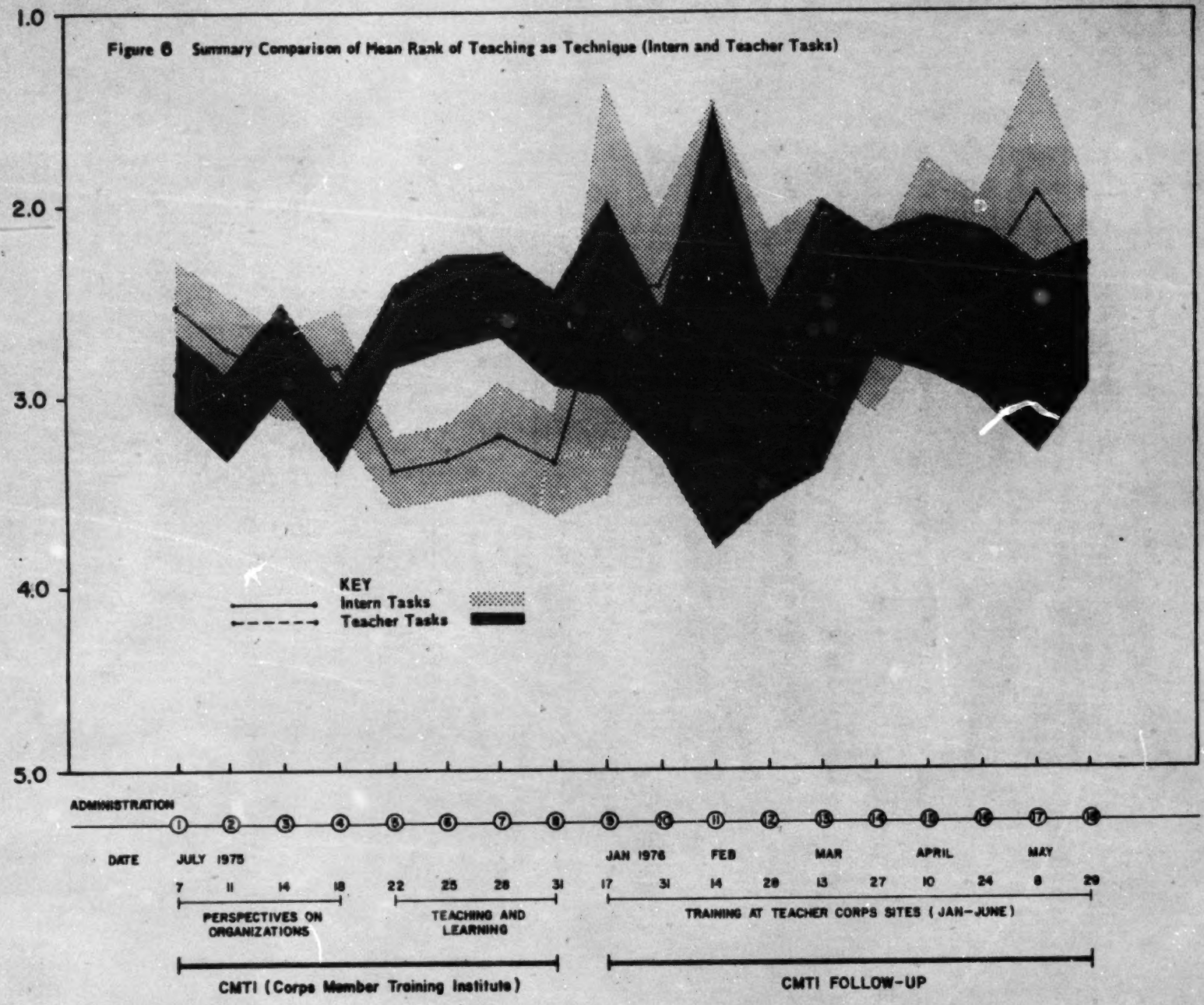
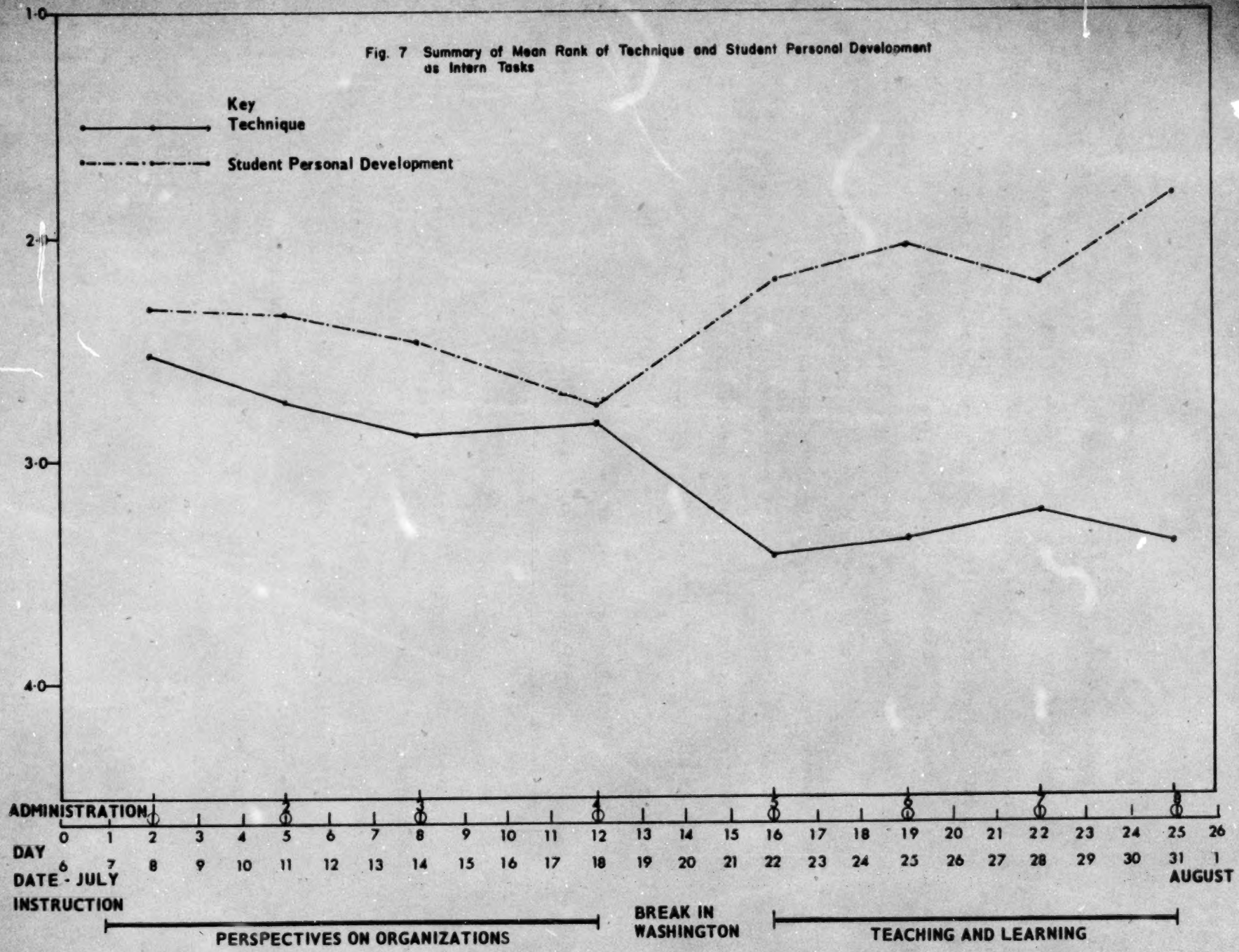


Fig. 7 Summary of Mean Rank of Technique and Student Personal Development as Intern Tasks



more we found the following to describe the relationships being drawn about teaching as technique and teaching as student personal development. First, many seemed to apply a type of means-ends relationship between these two perspectives. Some, for example, saw techniques as the goal of teaching and potential means for reaching these goals as student personal development. Others saw student personal development as the goal and techniques as the means for reaching those goals. The ambiguity of these findings were further reflected in our discussions with participants about the meanings they applied to these two perspectives.[7] Most important, however, was the certainty with which we could say the reflection about this natural dilemma in teaching (technique vs. student personal development) was never undertaken by the participants before we had started our questioning.

Thus, a final hypothesis became the following:

(iv) There is no professional perspective being developed by participants during the first year of training.

In short, the survey results, graphs, data analysis, interviews and observations along with the interpretive process and the continuous generation of hypotheses questioned our original assumption, that is, that the socialization process would be directed towards participants' perspectives of what teaching was. Instead, we concluded (continuing our interest in the socialization process over professional development) that, except for the one month training event, participants were being socialized into the adult world. The dilemmas that were being

addressed by the student teachers were dilemmas in adult-adult relationships, not in professional perspectives towards teaching.

Thus, the evidence from the quantitative and qualitative data gathered and the interpetive processes used to generate, review and regenerate hypotheses suggested the following conclusion about the impact of the training event.

> During the training event there was an immediate, almost unconscious impact upon the participants' perspectives towards teaching and their interpretation of where they fit into the teaching profession. This impact was due to the processes of training, certain underlying assumptions of the trainers and the rich interactive experiences of the participants. This immediate, dramatic and unexpected impact was never followed up by actions or by reflection, thus making it an independent event, powerful in its own right but almost meaningless in the year's activities following that event.

Figure 8 summarizes the generation of the original hypothesis to successive hypotheses and an eventual conclusion about the event and its impact upon the participants' development of professional perspectives. As is evident from this discussion, the time-series design and data analysis contributed to the generation of successive hypotheses.

Political Affiliations

This was an evaluation study and, as such, was considered to be an intervention upon the educational environment that we were describing and upon the participants with whom we were talking or sharing our tentative findings. As Ross and Cronbach (1976) and others have noted,[8] this is a natural feature of evaluation and one that helps define the evaluation problem, the actual approach to the problem and the evaluative results and recommendations. It may be especially appropriate for us to place

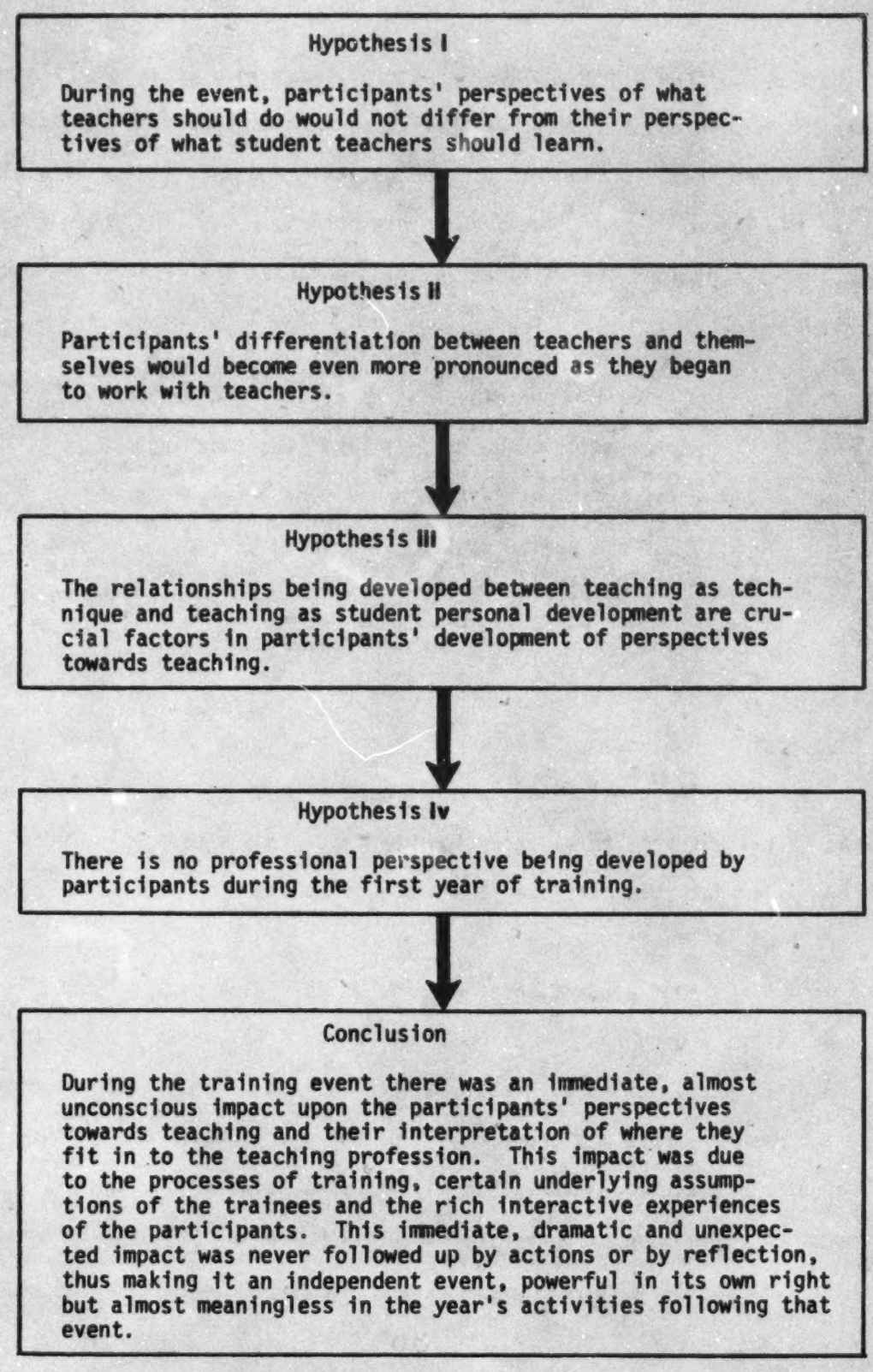

Hypothesis I
During the event, participants' perspectives of what teachers should do would not differ from their perspectives of what student teachers should learn.
Hypothesis II
Participants' differentiation between teachers and themselves would become even more pronounced as they began to work with teachers.
Hypothesis III
The relationships being developed between teaching as technique and teaching as student personal development are crucial factors in participants' development of perspectives towards teaching.
Hypothesis Iv
There is no professional perspective being developed by participants during the first year of training.
Conclusion
During the training event there was an immediate, almost unconscious impact upon the participants' perspectives towards teaching and their interpretation of where they fit in to the teaching profession. This impact was due to the processes of training, certain underlying assumptions of the trainees and the rich interactive experiences of the participants. This immediate, dramatic and unexpected impact was never followed up by actions or by reflection, thus making it an independent event, powerful in its own right but almost meaningless in the year's activities following that event.

Figure 8. Example of the Generation of Successive Hypotheses Used and a Conclusion on the Impact of the CMTI Training Event Upon Participants' Perspectives Towards Teaching.

this study into its political context since that, after all, is where the feasibility of our application of time-series design and analysis will be judged for future evaluation purposes.

There were four levels at which political relationships had to be maintained. One was, naturally enough, with the parent funding agency, Teacher Corps, who were sponsoring the training event and the evaluation study. Their concern was on a national level with the general need for (a) information on the event and its impact for help in determining future similar events and (b) information on the status and performance of participants who were being placed into an entirely different project environment than the agency had sponsored in the past. Since "demonstration" (a craft seen to be similar to evaluation) was also a new responsibility of the agency,[9] any insights from our study into unique possibilities for documentation, interpretation or communication of local project activities, successes and failures would also be appreciated and acknowledged as added benefits by the agency.

A second level of political relationships were with the instructors of the training event. Their concerns were professional, that is we were not only intruders upon their carefully designed learning environments, we were potentially destructive to their professional status. Here, concerns, although expressed in professional language about study design and analysis, were understandably personal as well.

A third level of political relationships were with the local project directors to whom the participants in the training event were returned. In a sense, they were in the most difficult position because (a) they had

many other responsibilities to carry out besides the training of four student teachers, (b) their relationships with the funding agency were often felt to be sensitive, (c) they already felt they "knew" what was going on at their own sites and felt that they never would really "know" what went on during that first month at the training event, and (d) in a sense, they were asked to cooperate with us more than anyone. Generally, they had little to gain from our study and much more to lose.

The fourth level was with the participants themselves. Their concerns were also both personal and professional. As beginners in a new profession, placed in new surroundings and as lowly members of a complex infrastructure, they were most interested in having a professional advocate and an open line of communication to "those in charge." Although our study was an intervention mostly upon their time and thoughts, generally they were most receptive to us, not necessarily because we were sympathetic but because we were, besides being listeners from their chosen profession, also an alternative line of communication from themselves to others in the agency. For example, their critique of our drafts would generally include not only (a) did it capture our feelings and actions but also (b) will it have an impact on what will be done in the future?[10]

As any evaluator knows, the interrelationships between these levels were, in a political sense, just as real as the people themselves. In other words, the relationship drawn between instructors and parent agency, between parent agency and local project or between participants and instructors were also political realities that would determine how an

evaluation team could negotiate their own relationships with these levels while adhering to a general study direction. We were given the general problem by the parent agency, told the context(s) for which they could use the information and then were left to further define the problem and carry out our responsibilities using our own professional interpretations. There were certain characteristics of the evaluation settings which made this "autocratic" approach possible.[11] We could deal with the instructors on an equal professional level partly because we, too, were teacher educators, partly because they shared some of our professional perspectives, partly because the combination of time-series and field study techniques interested them. The local project directors were generally too busy to care much what we did as long as we respected their logistical dilemmas and kept contact with them and their trainees. Most important, however, was that we very easily and naturally aligned ourselves to the strongest relationship among these four levels: that is the relationship between the student teacher participants and the parent agency.

Our acceptance of the general problems of the agency for policy relevant context information, our professional respect for the purpose of the event and what it could tell the profession and our advocate role for the participants made it possible to negotiate many difficult features of our study design. For example, the sticky problem of the number of survey administrations (especially during the training event) were eventually negotiated to be eight. Likewise, the opportunity to be responsive to both parent agency and participants made it possible to include

uncomplementary statements about the training event or the context of the local project training sites.[12] In short, the political context was real, was complex and was always dealt with in an openly political manner where potential confrontations were recognized and respected and where fortunate alliances made it possible to carry out the complexity of the study design.

The point that is being emphasized here is that we applied time-series design and data analysis into a typical evaluative setting where the political context in that setting played a role in how the study was designed, and carried out. In fact, in this political context, it was the requirement of the time-series design for qualitative data, and the underlying assumption that there were other processes being intervened upon, combined with the intriguing and provocative time-series results immediately following the training event, that contributed to the political feasibility of our evaluation design.

To summarize this discussion on how time-series design and data analysis was integrated into our program evaluation, it is important to understand their contribution to our dialogue. Dialogue, Mayer (1976) suggests, is cooperative inquiry aimed at:

(1) formulation of a problem which is not susceptible to convincing empirical demonstration,
(2) the statement of possible solutions and interpretations of the problem and
(3) methodological considerations of those solutions and interpretations in terms of both common and particular circumstances.

Mayer further suggests that it is an instrument (he says, "the instrument") for continuing investigations of matters which lie outside the realm of positive proof.

There is no doubt that many program evaluations often must deal with problems that are "not susceptible to convincing empirical evidence." In fact, in this sense of dialogue, the mode that we chose for program evaluation could rightfully be called "dialogue-evaluation" where the dialogue was not only between people of different perspective and persuasions but also between evidence that was qualitative and evidence that was quantitative. For this dialogue, information and perspectives supplied by time-series design and time-series analysis were valuable contributions.

Feasibility: The Advantages, Limitations and Difficulties in Applying Time-Series in Program Evaluation

The feasibility of time-series design or data analysis does not rest upon its cost or its technical nature. Both are minimal and are readily accessible to most evaluators. Instead, feasibility is determined by what may reasonably be expected to be done in a particular setting. The following is meant to suggest the feasibility of time-series design and analysis in program evaluation by discussing specific advantages and specific limitations of time-series from an evaluation perspective. Difficulties that we have encountered and that we believe are endemic to the application of time-series to most program evaluation challenges are also discussed as feasibility considerations. After potential feasibility of time-series has been presented in this way, there will be stated some suggestions for those who may be considering a time-series design or data analysis in the evaluation of an educational program.

Advantages of Time-Series Design and Analysis

There are certain features of a time-series design that are especially attractive and, sometimes, uniquely appropriate to educational program evaluation. The time-imbeddedness of a program intervention often not only reflects the reality of an educational intervention but also points up some of the salient features of a program. Time and timing are crucial factors in an educational program where events, critical incidents and work loads are often major context features that can be portrayed when time is a variable. The interpretation of the effect of an intervention as a disrupter of a process rather than as an "outcome" is most often the reality of an educational program as well as unique to a quantitative approach. Likewise, the variety of possible effects (e.g. delayed or temporary effects, changes in direction or level) are much more illuminative if real impact rather than statistical differences between pre and post measures is the major concern.

One of the most attractive features of the time-series design and analysis is that it is such a useful and meaningful way to combine quantitative and qualitative data. In addition, the results of this combination are capable of being nicely related and communicated through graphics (see Figures 4, 5, 6, 7). Furthermore, this graphic portrayal of context variables and the dependent variable make it a very attractive source of evidence and a rich motivator for heuristic interpretations from a variety of different perspectives. In the reports of our evaluation study, for example, we found the graphics of figures such as Figures 4, 5, 6, and 7 powerful and potent for communicating our results to agency policy makers, training instructors or participants. More important, not only did we find these

graphics cogently forceful, we also found they helped focus attention on some of the ambiguities of the data and helped the audience raise the most crucial question of what was happening when and to whom (both, traditionally, some of the more difficult perspectives for policy makers to take).

Moreover, one can conceptualize the usefulness and appropriateness of time-series design and analysis to a variety of different modes of evaluation. A time-series design, with multiple readings across time accompanied by descriptive statistics (e.g. Figure 7) may be quite valuable as the kind of direct, quick and informative evidence that is needed in formative evaluation. For summative evaluation, time-series design and analysis may give real insight into the impact of one program as well as associative information about the comparative, concomitant variations between the impacts of different programs. Certainly time-series design and analysis are consistent with the purposes, intentions and focus of discrepancy evaluations. Since time-series is a search for discrepancies (and similarities) in the patterns of a variable at different times, there are many questions that may be asked in discrepancy evaluation that may be uniquely illuminated by a time-series design and analysis.

Perhaps the most interesting feature of time-series design and analysis, however, is its potential to a variety of evaluation modes that have traditionally questioned the usefulness of quantitative data. Illuminative evaluation (Parlett and Hamilton, 1976), for example, seeks continuous interpretations, a longitudinal perspective of a program and a focus on

the day to day reality of a program, all features of a time-series design and analysis. The directness of time-series results and its sensitivity to practitioners' awareness that most phenomena they are interested in are dynamic and changing make time-series design and analysis potentially attractive to responsive evaluation (Stake, 1974). Likewise, the dynamic and changing patterns that can be highlighted by time-series design and analysis may be especially illuminative to ethnographic evaluations (Carini, 1975). Highlighted patterns in a variable such as "time-on-task", for example, could illuminate certain customs that are being tacitly followed in an educational program. Thus, time-series design and analysis may not only be potentially attractive to evaluation modes that have traditionally used quantitative data (e.g. summative or discrepancy evaluations), it may also be appropriate for more qualitative modes of evaluation such as responsive, illuminative or ethnographic evaluation.

To summarize the advantages of time-series design and analysis: it offers the possibility for evaluators to anticipate meaningful fluctuations in the observed variable, to note the imbeddedness of the setting in other time and context frameworks, to interpret and describe the intervention as a dynamic and changing event and to use a variety of perspectives for inferring both intended and unintended effects. In short, time-series design and analysis can be an interpreter's dream: rich in evidence, provocative in perspective and ambiguous or suggestive enough for creative interpretation and real dialogue.

Limitations of Time-Series Design and Analysis

The limitations of time-series design and analysis for program evaluation in educational settings are just as real as the advantages. The number

of necessary observations are usually the first serious concern. In educational evaluation, the concern is not only for test-wiseness (if the data is collected through a survey, test or questionnaire) but it is also political. Who will allow sixty (60!) readings across time? (Suggested by Box and Jenkins, 1970, for economics.) What we have found is four things. One, it is possible to negotiate for more readings than had been imagined (e.g. eight during a four week event). Two, it is possible to use matrix sampling to help address the test-wiseness problem in an evaluation with limited population. Three, the relationships drawn between us and the participants as part of the descriptive responsibilities of the time-series design helped us immeasurably with participant response. Four, a time-series analysis minimally performed with about twenty observation points can still be provocative. For example, we had reasonably good results with just eight data points during the training event (see Figure 7). Although these were not analyzed at that time by time-series analysis, the descriptive data (means, variances and confidence intervals) along with the unexpected changes were provocative and informative to both the evaluators and to the agency policy makers. Thus, from our experience, although the number of observations is a real problem, it is not necessarily the greatest problem that an evaluator will face when applying a time series design.

A second limitation is the need for a data base before the intervention begins in order to perform a time-series analysis. In economics, a suggested minimum number of pre-intervention scores is twenty, in the Fortran program (Bower et al, 1974) an absolute minimum is three. We performed

time-series with no actual pre-intervention scores. To our study, this meant that our choice of the two intervention intervals were arbitrary and not congruent with the actual intervention (i.e. the training event). This made interpretation of the time-series analysis a bit more complicated than it would have been if the actual interventions were congruent with the intervals chosen for analysis. Furthermore, the results of our study were so unexpectedly dynamic both during and after the actual intervention, that it was unfortunate that we did not have a real pre-intervention data base to compare these results. Thus, although the lack of a pre-intervention data base was unavoidable (i.e. many participants were not identified until a day or two before the training event) it was also unfortunate. In fact, even considering the excitement of working in a multidisciplinary setting (which was a major motivating factor for choosing time-series design) we suspect that we would not attempt applying a time-series design and data analysis to future program evaluations unless there was the possibility of collecting at least three actual pre-intervention observations.

A third limitation is the requirement for univariate dependent measures. Although, as has been noted, this problem is currently being worked on by Swaminathan and Algina (1976) and Geweke (1977), and progress in making time-series analysis multivariate may be occurring at a rapid rate, it would still be unwise to suggest it is accessible for application in most evaluation studies. Thus, the univariate character of dependent variables in a time-series analysis either makes the choice of variables more limited or the interpretation more heuristic. We, for example, with the use of

ranking scores, chose the latter course, where the time-series design and analysis contributed to the generation of hypotheses and the evolving dialogue.

A fourth limitation, and one shared by most quantitative methodologies, is the requirement for perfectly reliable data. Since fluctuations in the dependent variables are assumed to be real as well as important when analyzing a particular series of that dependent variable, it is important to have stable measures with no large unaccountable sources of variation. Time-series analysis must have highly reliable data points in order that fluctuations due to measurement errors are not interpreted as actual shifts or changes in level of the dependent variable. In our study the reliability of the surveys used were high (.82 to .92, using Hoyt's analysis of variance). In this case as in most educational situations, one assumes that the measurement errors are randomly distributed and will not affect the major changes in level or changes in drift determined through the time-series analysis. Another perspective taken in this study toward checking the reliability of the survey data was to record the confidence intervals for all mean scores. Although reliability is not necessarily related to confidence intervals, it can be argued that the confidence intervals themselves can be used as reliable predictors of group means.

Finally, it must be recognized that, although unexpected events may be charted and dynamic fluctuations of a variable may be noted, nevertheless, once a dependent variable is chosen and the study has begun that choice cannot be changed. To some perspectives on evaluation, this a-priori feature of time-series is attractive but to other perspectives the inflexibility of the choice of dependent variables is unfortunate.

As was described earlier, although the dependent measures in our study remained the same, some of the connotations and meanings of those variables changed (along with our hypotheses) as we noted changes or became more acquainted with the participants' views.

Note that there are some characteristics of time-series design and analysis that are advantages from one point of view and weaknesses from another. The inflexibility of the choice of measured variables, for example, is one feature that can be interpreted as an advantage, from one perspective, but as an unfortunate limitation from another. Similarly, the heuristic interpretive features of time-series data analysis may be seen as an attractive feature from one perspective of evaluation (e.g. illuminative evaluation) but may be considered as limiting its usefulness to another perspective (e.g. discrepancy evaluation).

Difficulties of Time-Series Design and Analsis

In addition to noting certain advantages and limitations of time-series design and analysis, it may be well to point out other features of time-series that are simply difficulties that need be addressed. They are not necessarily an advantage or a limitation but instead are accompanying features to the study design and data analysis that must be given thought and consideration when applying time-series to an evaluative setting.

We learned, for example, that it is necessary to interpret the time-series data periodically during the evaluation study. Although this is a relatively minor difficulty that can be overcome by either descriptive analysis or time-series analysis, it suggests that data should be transcribed, plotted and analyzed at a time when other time-consuming features of the study are occurring (e.g. interviews and observations). In our study,

although we performed a major documentation and review of the data at the end of the training event, we did not review the data again until late in the follow-up phase of our study. This was unfortunate, since the periodic review sessions and evolving focus of the follow-up phase could have been further illuminated by the quantitative results from the time-series.

A second difficulty that is often a feature of a time-series design and analysis is the ambiguity of the results and its potential for over-interpretation (see Glass et al). Ambiguity is, in part, related to the number of data points where analysis represents compromises between the data points and human behavior. Overinterpretation is the other side of the coin when recognizing the appeal and the directness of the graphic representation of the results. The charts and results from the event phase of our study (see Figure 5) did serve the interpretive needs of the study and formed a basis for dialogue and redirection. The ambiguities of the results in later administrations, however, have made our interpretations of impact or effect over time much more difficult. Time-series analysis, of course, reduces the ambiguity (the more data points, the less the ambiguity), but it is here that it becomes difficult to combine the quantitative graphic portrayal with quantitative analysis and with qualitative descriptions and interpretations. The results depicted in Figure 9 are a very good example of the dialectics thay may occur between the portrayal, the analysis and the descriptions of the context.

Figure 9 shows the rankings given to community involvement as an intern task, that is, the relative importance of using community resources

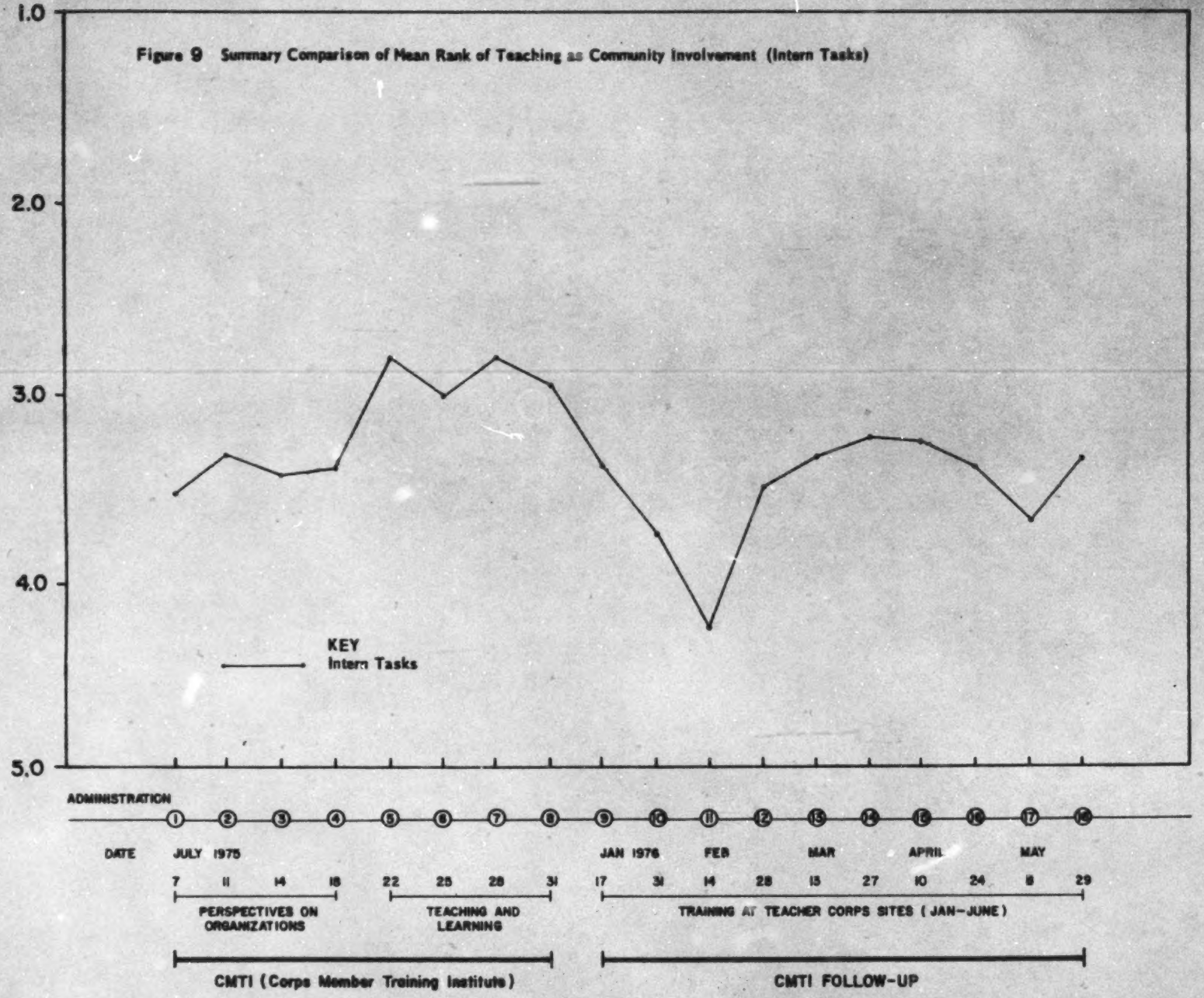


Figure 9 Summary Comparison of Mean Rank of Teaching as Community Involvement (Intern Tasks)

in the school and providing services to the community. The direct interpretation for the graphics of the first eight administrations (shown also in Figure 5) is reduced considerably by the dynamics of the results in the last ten administrations. Time-series analysis helped us interpret these findings, suggesting that the ranking of community involvement was depressed during the follow-up phase. This analysis was very interesting in light of our findings that community involvement composed the major responsibility of the participants during the follow-up phase.[13] Instead of placing the most meaning upon this their major field of responsibility, the participants reacted to an obvious lack of status of this involvement in the eyes of others and correspondingly rated community involvement progressively lower as this became more apparent. This interpretation of gradual depression, from the time-series analysis, however, does not conform entirely to the graphic results. The extreme depression noted in administration number 11, for example, was not highlighted in the time-series analysis. Correspondingly, it was not addressed by us (except as an aberration) until we were reminded by participants that that time was when they were first graded as student teachers! It was at that time, they told us, that the unimportance of their time-consuming activities of community involvement became most apparent. Here, then, is an example, when ambiguity, overinterpretation and a dialectic between the data points, the analysis and the description of the context all were necessary to suggest meaningful interpretations and locate critical incidents. Difficult results like these may be more expected from a time-series design than the dramatic, directly interpretable results noted in the event phase of the study (e.g. Figure 5 and 7).

The most serious difficulty, however, is the requirement for real dialogue between quantifiers and qualifiers of experience. To say that the data are dialectically related is one thing. To get dialogue to occur between those who are most qualified in interpreting these two kinds of data is an entirely different matter. First, it must be realistically noted that this dialogue is a rarity (see Mayer, 1976), making the responsibilities for dialogue more difficult to carry out. The amount of learning about dialogue and about each other that need be accomplished by team members should not be underestimated. The importance of the myths that need be dispelled cannot be overlooked (e.g. qualifiers thinking that quantity is only Newtonian mechanics, quantifiers thinking that qualitative procedures produce unreliable and useless data). An increase in dialogue between qualifiers and quantifiers of human experience can only occur through great effort, repeated confrontations and common concerns. There were, for example, some common concerns shared by members of our study team that were crucial to the generation of any dialogue between us. These common concerns were towards: (a) the time-imbeddedness of the event we were evaluating, (b) the potential relationships between training context and dynamic changes in professional perspective, (c) the identification of underlying characteristics of the training event that would place it in a larger context (e.g. perspectives on teaching) and (d) the potential of a socialization metaphor for describing teacher professional development. The paradox is that dialogue requires both strong differences in professional perspective as well as common concerns. The most difficult aspect in approaching this paradox is the relative non-existence of dialogue in professional life.

Summary: Feasibility of Time-Series Design and Analysis

It is not the cost of time-series either as a design (e.g. number of surveys given) or as a data analytic process (e.g. Fortran IV computer package) that determines the feasibility of time-series for program evaluation. Both were minimal costs of this study (less than 5% of study expenses). Neither is it the requirement for expertise of experience in time-series analysis (we had neither when we started and found the available literature explanatory). Instead, the feasibility of a time-series design and data analysis for a particular evaluation study is determined by the time available for planning, dialogue and interpretation, the persons available for quantitative and qualitative data gathering and data analysis, the amount of interaction that can be expected between these persons, the importance and value of dialogue to these persons, the relationships drawn between various patrons, clients and participants, the interest and fascination with the evaluation problem shared by members of the evaluation team, the shared assumptions about the nature of educational environments and the corresponding challenges for inquiry, and, especially if it occurs in a university setting, the willingness of the director to negotiate and allow for individual autonomy as well as team dialogue and production. Of these considerations, the most important feature that determines the feasibility of using time-series design and analysis is the extent to which real dialogue can be anticipated and supported.

Suggestions for the Application of Time-Series Design and Analysis to Program Evaluation

For those who are still considering applying time-series to program evaluation, even after our discussion of the limitations and difficulties of time-series design and analysis, the following suggestions are offered. These are offered to help those who may choose a time-series design and analysis get through some of the difficulties not addressed by either Glass, Willson and Gottman (1975) or Campbell (1976). Most important, however, is that these guidelines are offered so that the full potential of an adaptation of time-series design and data analysis to educational program evaluation can be realized.

1. Time-series design and analysis may be especially valuable to those anticipating a dynamic, innovative event where some of the "limitations" and "difficulties" of time-series (e.g. the ambiguities and requirements for heuristic interpretation) become real advantages when considering the evaluation challenge. Thus, we are suggesting time-series design and analysis may by most applicable to those modes of evaluation (e.g. illuminative, ethnographic, responsive, democratic) that have traditionally shunned quantitative methodologies.

2. Use time-series design and analysis for time bounded events that cannot be replicated (but possibly be adapted or retried) and where dynamic happenings may by expected but one is not sure where or when critical incidents will occur.

3. Use a team approach in the evaluation design, leaving plenty of lead time for planning and conceptualizing the problem together. A variety of cultural, professional and institutional role perspectives is valuable. Realize that involvement in the evaluation study may have an impact on the evaluation team members.

4. Build in many review sessions, periodic reports and regular periods for analysis and interpretation of collected data. One must plan for dialogue by having day-long sessions rather than short-term hourly meetings.

5. Choose variables that are theoretically inspired and can help interpret the imbeddedness of the intervention as well as illuminate unexpected outcomes.

6. Keep clients (and patrons) involved at all times. Communicate one's evaluation intent, how to "read" the results and interpretations, the theoretical basis (and bias) of the study, evolving surprises of the study and ambiguous results that have been interpreted. In short, time-series design suggests a different approach to clients as well as to the phenomenon being evaluated.

7. Be prepared for illumination rather than conclusion. Given the lack of information available from a longitudinal perspective, we are simply not prepared for the surprises that occur when our assumptions about linearity are questioned.

8. Provide as much context information as possible. This means working closely with people from two different highly trained crafts: quantitative and qualitative investigation.

9. Try to get individual data if it is possible. We did not have data on individual persons that could be analyzed through time-series analysis. It could have provided us with some interesting questions and potential answers.

10. Report through graphics, photography and informal presentations, as well as through writing. In fact, it is the writing that is one of the major problems we encountered (especially of the "final" report). We found it extremely difficult to combine our perspectives and styles and have a unified final report. There are two possibilities here. One, have one writer whose drafts are reviewed by all members of the team. Two, have one interpretation written with other interpretations printed as responses, clarifications and challenges to the given interpretation (this was suggested by John Elliott). We even found a poet helpful.[14]

11. Employ a graphic artist and a photographer. The choice of graphic itself becomes a professional interpretation of the results and can be a powerful communicative instrument. For example, we found photography helped us analyze some of the more difficult, imbedded features of the interactions that occurred during the training event.

12. Provide as many perspectives on an event as possible for descriptions and interpretations of what happened and why. In our study, perspectives from curricular theory, professional socialization, organizational theory and teacher education all helped illuminate our understanding of the event and its impact.

In conclusion, program evaluation, as Lee and Cronbach (1976) have recently suggested and as MacDonald (1974) and Apple (1973) have continually argued, is a different craft from research. Some of the problems encountered in evaluation are unique and some of the criteria used to judge the potential effectiveness or appropriateness of a suggested methodology are special to evaluation. Many of the unique problems in educational program evaluation, such as programs being time-bounded, context specific, dynamic and uncontrolled environments not always capable of convincing empirical evidence alone are just those problems that make time-series design and analysis so appropriate for program evaluations and so valuable to their purposes.

Footnotes

1. A concise introduction to Teacher Corps and its recent changes in direction can be found in the report of the National Advisory Council on Education Professions Development, titled Teacher Corps: Past or Prologue?, 1975.

2. The contract was with Teacher Corps Recruitment and Technical Resources Center, University of Nebraska-Omaha, contract no. 300-75-0100.

3. Principal investigators of this were: G. Thomas Fox, Jr. (director), Carl Grant, Thomas Popkewitz, Thomas Romberg, B. Robert Tabachnick and Gary Wehlage at the University of Wisconsin-Madison, School of Education, Department of Curriculum and Instruction.

4. Parlett and Hamilton (1976) and Hernandez-Nieto (1976) provide related perspectives on this approach to evaluation.

5. Technical Report Number 4, "Using Photographs for Interpretation", J. Tabachnick and G.T. Fox in Fox et al, 1976 demonstrates how evaluation can be used for analysis. Becker (1975) discusses the general value of photography as a sociological perspective. In our study, Becker's suggestion that graphic metaphors may be uniquely supplied by photography was reflected in our choice of the phrase "a sense of community" which had graphic reference whenever we used that term to describe the training event.

6. This reflected findings and interpretations of earlier studies, see Fox (1976) and Teacher Corps: Past or Prologue? (1975).

7. A card sort was conducted with the participants with the results suggesting much confusion between the categories "student personal development" and "technique". See Technical Report Number 6, "CMTI Impact Study Design and Survey Data Analysis" in Fox et al, 1976, pates 21-23.

8. Gary Wehlage in Technical Report Number 21, "The Ethics and Politics of Evaluation: Clients, Patrons and Casualties", raises many of the political dilemmas encountered when the patrons (e.g. the agency) and the clients (e.g. the instructors) are not the same parties.

9. As is suggested in Teacher Corps: Past or Prologue? (1975) or in the Federal Register Volume 40, No. 243, Wednesday, December 17, 1975., pp. 17-43, Teacher Corps projects became demonstration projects of innovative teacher training strategies.

10. To the participants, there was always a tension between us as the outside, impersonal "testers" and the personal, concerned "listeners". This odd relationship between the participants and ourselves never ceased and, in many respects, defined the mutual respect we had for each other.

11. Barry MacDonald (1974) categorizes evaluation into three political categories: "autocratic", "bureaucratic" and "democratic". He probably would have labeled our study "autocratic". Although we did more sharing and negotiating than that term was meant to imply, we were given the autonomy to pretty much define the problems and our approach to the problems.

12. In fact, the most uncomplementary report was received most warmly (and had the greatest impact) by the national agency and by the local directors. This was Technical Report Number 13 in Fox et al (1976) titled "Alternative Solutions to the Problems of Training Eleventh Cycle Teacher Corps Interns." This report was shared in the spring of 1976, well before the technical reports were published.

13. See the description of the follow-up contract titled "Description of Internship" by T. Popkewitz, G. Wehlage and C. Grant, Technical Report Number 5 in Fox et al, 1976.

14. See Technical Report Number 3, "Poem" by Wendy Saul, in Fox et al, 1976.

REFERENCES

1. S. Anderson, S. Ball, R. Murphy and Associates, Encyclopedia of Educational Evaluation, San Francisco: Lossey-Bass Publishers, 1975.

2. T.W. Anderson, The Statistical Analysis of Time-Series, New York: John Wiley and Sons, Inc., 1971.

3. Michael Apple, "The Process and Ideology of Valuing in Educational Settings", Educational Evaluation: Analysis and Responsibility, Apple, Subkoviak and Lufler, (eds.), Berkeley: McCutchen Publishing Corp., 1974.

4. Howard Becker, "Photography and Sociology," Afterimage, 1975, May-June, 21-32.

5. I. Bernstein and H. Freeman, Academic and Entrepreneurial Research, New York: Russell Sage Foundation, 1975.

6. B. Bloom, J. Hastings and G. Modans, Handbook on Formative and Summative Evaluation of Student Learning, New York: McGraw-Hill, 1971.

7. C. Bower, W. Padia and G. Glass, TMS: Two Fortran IV Programs for Analysis of Time-Series Experiments, Boulder, Colorado: Laboratory of Educational Research, October, 1974.

8. G.E. Box and G.M. Jenkins, Time Series Analysis, Forecasting and Control, San Francisco, CA: Holden-Day, 1970.

9. A. Bussis, E. Chittenden, M. Amarel, Beyond Surface Curriculum, Boulder, Colorado: Westview Press, 1976.

10. Donald T. Campbell, "Reforms as Experiments", Handbook of Evaluation Research, Volume 1, E. Struening and M. Guttentag, (eds.), Beverly Hills: Sage Publications, Inc., 1975, 71-100.

11. D.T. Campbell and J.C. Stanley, Experimental and Quasi-Experimental Designs for Research, Chicago: Rand McNally College Publishing Co., 1966.

12. P.F. Carini, Observation and Description: An Alternative Methodology for the Investigation of Human Phenomena, Grand Forks, ND: North Dakota Study Group on Evaluation, University of North Dakota, February 1975.

13. A. Clegg, L. Anglin, R. Alfonso, A. Hammons and E. Migia, "Field Based Teacher Education: Management Research and Evaluation", Symposium presented at AACTE Convention, March, 1977.

14. David Cohen and Michael S. Garet, "Reforming Educational Policy with Applied Social Research", Harvard Educational Review, 1975, 45/1, 17-43.

15. Lee Cronbach, "The Two Disciplines of Scientific Psychology", American Psychologist, 1957, 12, 671-684.

16. Lee Cronbach, "Beyond the Two Disciplines of Scientific Psychology", American Psychologist, 1975, 30, 116-127.

17. E. Eisner, "Emerging Models for Educational Evaluation", School Review, 1971, 80, 573-590.

18. Elliott Eisner, "The Perceptive Eye: Toward the Reformation of Educational Evaluation", Stanford Evaluation Consortium, Department of Education, Stanford University, Occasional Paper, December 1975.

19. Federal Register, Wednesday, December 17, 1975, 40/243, 17-43.

20. G.T. Fox (Director), Impact of CMTI, Description and Preliminary Analysis, Madison, Wisconsin: Part Two of Interim Report, 1975 CMTI Impact Study, University of Wisconsin, October, 1975.

21. G. Thomas Fox, Jr., "Limitations of Standard Evaluation Methodologies for Federal Program Evaluations", Paper presented at 1976 AERA Convention, San Francisco, April, 1976.

22. G.T. Fox and R. Hernandez-Nieto, "Why Not Quantitative Methodologies to Illuminate Dialectical or Phenomenological Perspectives?", Paper presented at 1977 AERA Convention, New York City, April, 1977.

23. G.T. Fox, C. Grant, T. Popkewitz, T. Romberg, B. Tabachnick and G. Wehlage, 1975 CMTI Impact Study Technical Reports Numbers 1-21, Madison, Wisconsin: University of Wisconsin-Madison, September, 1976. Prepared as part of Teacher Corps, Office of Education Contract No. 300-75-0100 (University of Nebraska-Omaha).

24. John Geweke, "The Dynamic Factor Analysis of Economic Time-Series Models", Madison: Social Systems Research Institute, University of Wisconsin-Madison, Paper no. 7602, 1977.

25. G.V. Glass, "Analysis of Data on the Connecticut Speeding Crackdown as a Time-Series Quasi-Experiment", Law and Soc. Rev., 1968, 3, 55-76.

26. Gene Glass (ed.), Evaluation Studies Review Annual, Volume I, Beverly Hills: Sage Publications, 1976.

27. G.V. Glass, V.L. Willson and J.M. Gottman, Design and Analysis of Time-Series Experiments, Boulder, Colorado: Colorado Associated University Press, 1975.

28. Gene Glass, "A Paradox About Excellence of Schools and the People in Them", Educational Researcher, March, 1975, 9-13.

29. J.M. Gottman and R.E. Clasen, Evaluation in Education. A Practitioner's Guide, Itaska, Illinois: F.E. Peacock Publishers, Inc., 1972.

30. J.M. Gottman and R.M. McFall,"Self-monitoring Effects in a Program for Potential High School Dropouts: A Time Series Analysis", Journal of Counseling and Clinical Psychology, 1972, 39, 275-281.

31. E. Guba and D. Stufflebeam, Evaluation. The Process of Stimulating, Aiding and Abetting Insightful Action, Monograph Series in Reading Education No. 1, Bloomington: Indiana University, 1970.

32. David Hamilton, "Some Contrasting Assumptions About Case Study Research and Survey Analysis", To appear in Helen Simons (ed.), Towards a Science of the Singular.

33. G.E. Hein, An Open Perspective on Evaluation, Grand Forks, ND: North Dakota Study Group on Evaluation, University of North Dakota, February, 1975.

34. Rafael Hernandez-Nieto, Process-Development: An Alternative Approach for Curriculum Research in Mathematics, Madison, Wisconsin: Wisconsin Research and Development Center for Cognitive Learning, Working Paper no. 197, November, 1976.

35. Ernest House, "Justice in Evaluation", Evaluation Studies Review Annual, Volume 1, G. Glass (ed.), Beverly Hills: Sage Publications, 1976, 75-100.

36. Barry MacDonald, "Evaluation and the Control of Education", Innovation, Evaluation, Research and the Problem of Control, B. MacDonald and R. Walker (eds.), East Anglia, England: SAFARI PROJECT Interim Papers, Centre for Applied Research in Education, 1974, 7-19.

37. Barry MacDonald, "The Portrayal of Persons as Evaluation Data", Paper presented at 1976 AERA Convention, San Francisco, April, 1976.

38. Barry MacDonald and Malcolm Parlett, "Re-thinking Evaluation: Notes from the Cambridge Conference", Cambridge Journal of Education, Easter Term 1973, 74-82.

39. Milton Mayer, "Demise of Dialogue", The Center Magazine, January/February, 1977, Vol. X, No. 1, 12-15.

40. M. Parlett and D. Hamilton, "Evaluation as Illumination: A New Approach to the Study of Innovatory Programs", Evaluation Studies Review Annual, Volume I, Gene Glass (ed.), Beverly Hills: Sage Publications, 1976, 140-157.

41. M.A. Patton, Alternative Evaluation Research Paradigm, Grand Forks, ND: North Dakota Study Group on Evaluation, University of North Dakota, February, 1975.

42. M. Provus, Discrepancy Evaluation. Berkeley: McCutchen Publishing Corp., 1971.

43. Lee Ross and Lee Cronbach (eds.), "Handbook of Evaluation Research, Essay Review", Educational Researcher, November, 1976, 5/10, 9-19.

44. Michael Scriven, "The Methodology of Evaluation", Perspectives of Curriculum Evaluation, AERA Monograph Series of Curriculum Evaluation, No. 1, Chicago: Rand McNally, 1967, 39-82.

45. Michael Scriven, "Pros and Cons about Goal Free Evaluation", Evaluation Comment, 1972, 3/4, 1-14.

46. Michael Scriven, "Maximizing the Power of Causal Investigations: The Modus Operandi Method", Evaluation Studies Review Annual, Volume 1, G. Glass (ed.), Beverly Hills: Sage Publications, 1976, 101-139.

47. Robert Stake, "The Countenance of Educational Evaluation", Teachers College Record, 1967, 68, 523-540.

48. Robert Stake, "An Approach to the Evaluation of Instructional Programs (Portrayal vs. Analysis)", Paper presented at 1972 AERA Convention, Chicago, April, 1972.

49. Robert Stake, "Measuring What Learners Learn", School Evaluation, E. House (ed.), Berkeley: McCutchen Publishing Corp., 1973, 193-223.

50. Robert Stake, "Responsive Evaluation", New Trends in Evaluation, Institute of Evaluation, University of Goteborg, January, 1974, No. 35.

51. Robert Stake, The Measuring of Education, Berkeley: McCutchen Corp., (in press).

52. D. Stufflebeam, W. Foley, W. Gephart, E. Guba, R. Hammond, H. Merriman and M. Provus, Educational Evaluation and Decision Making, Bloomington: Phi Delta Kappa National Study Committee on Education, 1971.

53. E. Suchman, Evaluation Research: Principles and Practice in Public Service and Social Action Programs, New York: Russell Sage Foundation, 1967.

54. Hari Swaminathan and James Algina, "Multivariate Procedures for the Analysis of Quasi-Experimental Time-Series Designs", Paper delivered at 1976 AERA Convention, April, 1976.

55. J. Tabachnick and B. Lemes, Together: A Photographic Essay, Madison, Wisconsin: Part One of Interim Report, 1975 CMTI Impact Study, University of Wisconsin, September, 1975.

56. Teacher Corps: Past or Prologue? A Report with Recommendations, by the National Advisory Council on Education Professions Development, Walter Tice, Chairman, George Arnstein, Executive Director, Washington, D.C., July, 1975.

57. Rob Walker, "Classroom Research: A View from Safari", Innovation, Evaluation and the Problems of Control, B. MacDonald and R. Walker, (eds.), East Anglia, England: SAFARI PROJECT Interim Papers, Centre for Applied Research in Education, 1974, 20-24.

58. Rob Walker, "The Conduct of Educational Case Study: Ethics, Theory and Procedures", Innovation, Evaluation, Research and the Problem of Control, B. MacDonald and R. Walker (eds.), East Anglia, England: SAFARI PROJECT Interim Papers, Centre for Applied Research in Education, 1974, 67-101.

59. Carol Weiss, "The Politicization of Evaluation Research", Journal of Social Issues, 1970, 6/4, 57-68.